ESnet6 HighTouch Collector: Overview and Future

Kyle A. Simpson
k.simpson.1@research.gla.ac.uk
FelixMcFelix https://mcfelix.me

29th August, 2019

University of Glasgow

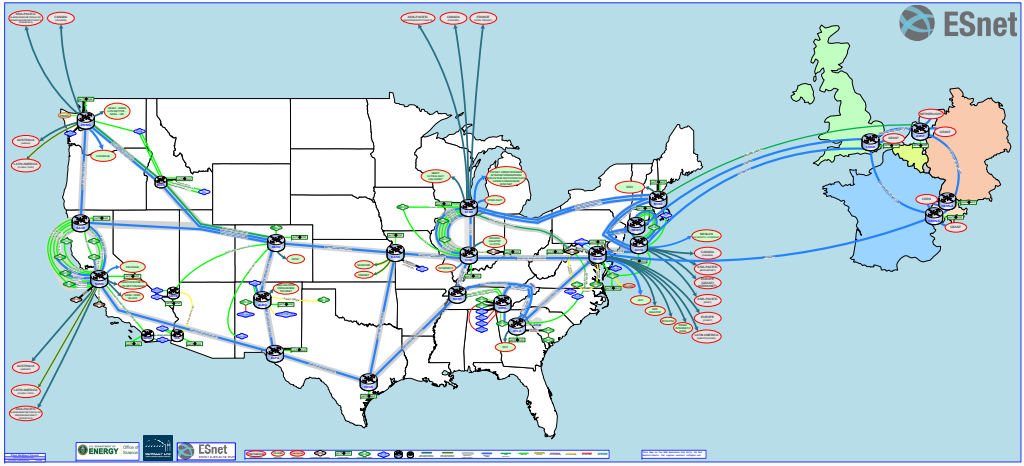
Introduction

Briefly, what is the collector? What can it do?

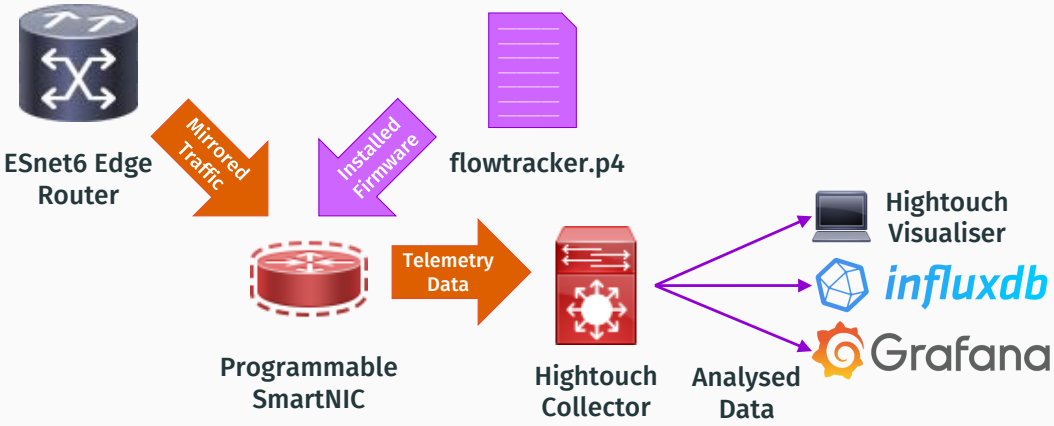
- Measure flow performance metrics.
- Stateful TCP analysis.
- Acts on every packet.
- *Fine-grained, line-rate.*

Why?

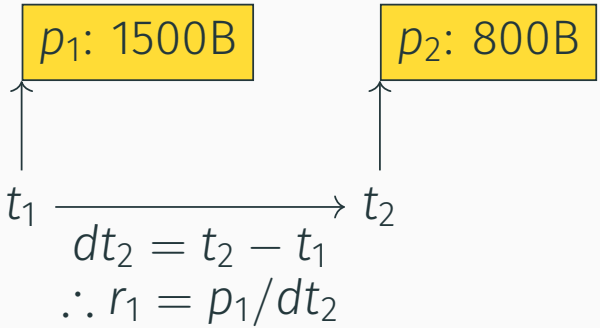
Large Networks



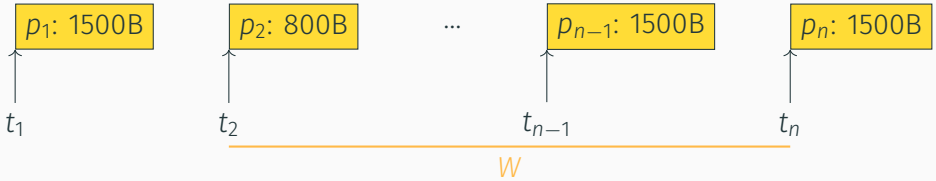
Where does the collector fit in?



Looking through the microscope...



...and zooming out.



Look over a *sliding window*, size e.g., $W = n - 2$.

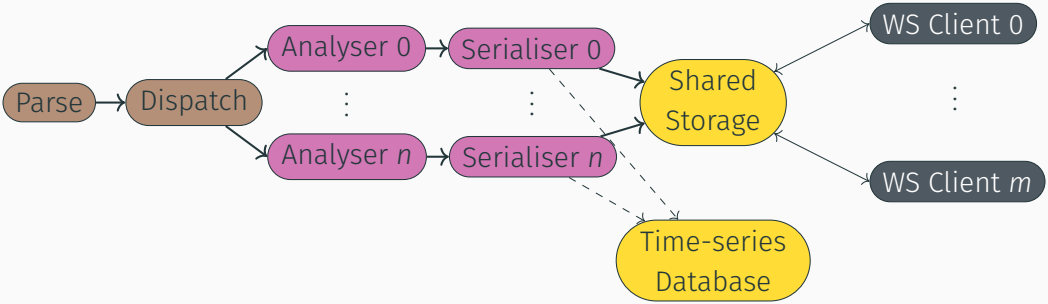
$$R_n = \frac{\sum_{a=n-W}^{n-1} p_a}{t_n - t_{n-W}}.$$

How is the Collector designed?

How does it operate?

- INPUT: Telemetry packets from SmartNICs.
- OUTPUT: Live time series of analysed data.
- OUTPUT: Mid-term storage of analysed data (time-series database).

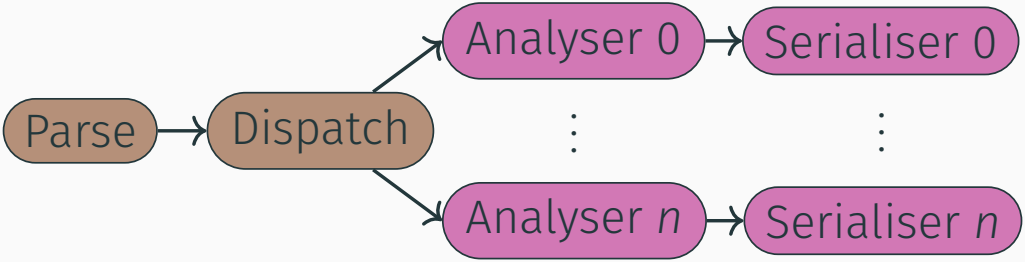
The Pipeline



Why design like this?

- One thread per stage $\implies$ increases throughput.
- More pipelines $\implies$ more flows at max throughput.

Performance



1300 kpps 775 kpps $n \times [310, 485]$ kpps $n \times 675$ kpps

63.5 Gbit s^{-1} Total 22.3–34.9 Gbit s^{-1} Per Flow

Analysis and Algorithms

- Rate monitoring (point vs. sliding window).
- Retransmission and loss detection.
- Initial SRTT estimation.
- Online half-SRTT estimation[1].
- Bytes-in-flight.
- Congestion window estimation[2].

[1] Karn and Partridge, 'Improving round-trip time estimates in reliable transport protocols'.
[2] Ghasemi, Benson and Rexford, 'Dapper: Data Plane Performance Diagnosis of TCP'.

Serialisation

- Each pipeline has a dedicated writing space, guarded by an RWLock.
- New messages (only in large batches) are written to this space, with timestamps.
- Any WS sockets iterate over all, take read lock, copy new messages… Then sleep on a condition variable at the top.

Instrumentation

So, what cool things can we see?

How do normal flows look?

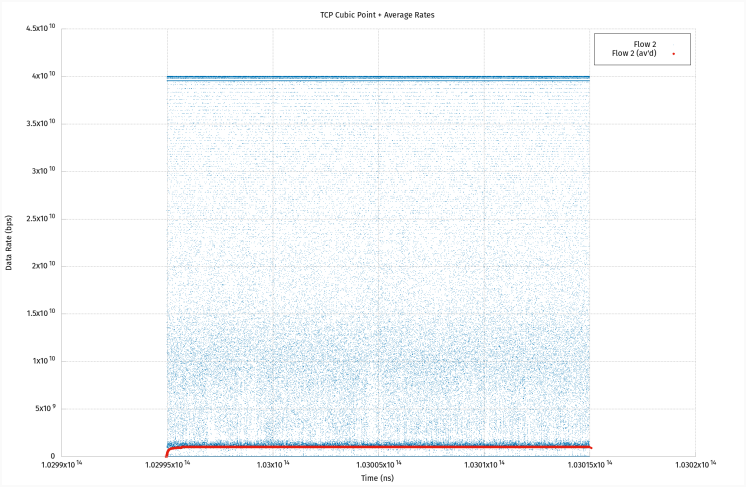


Figure 1: *TCP Cubic, 1Gbps*

Lossy flows?

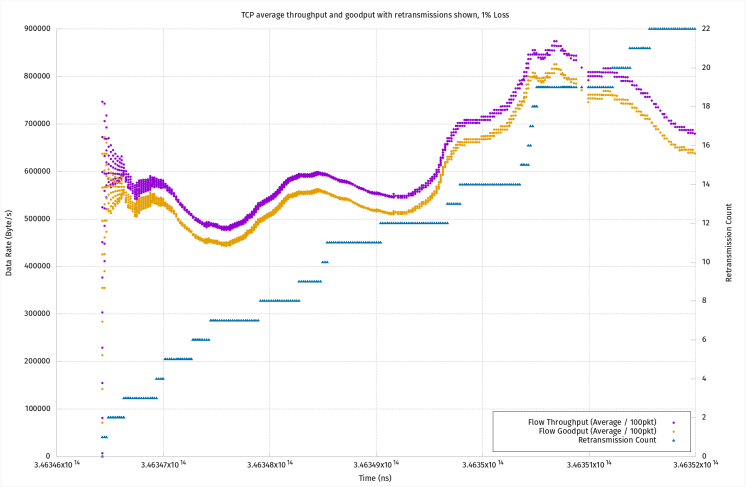


Figure 2: *TCP Cubic, [Intended] 1Gbps, 1% loss*

Multiplexed flows?

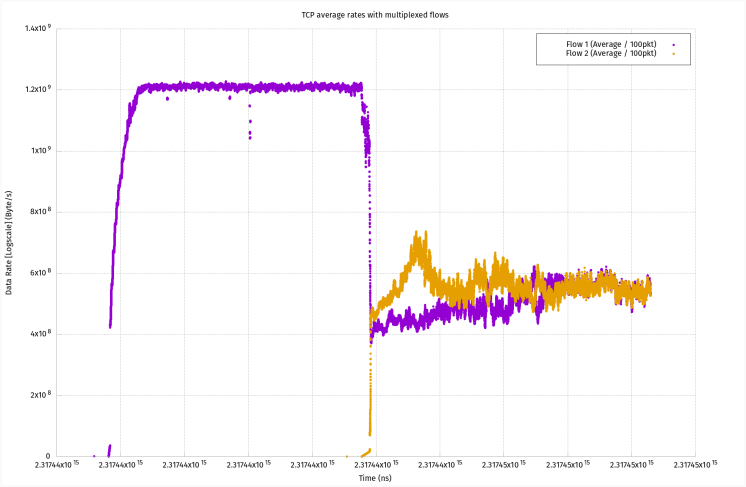


Figure 3: *TCP Cubic, 10Gbps, 2 flows*

Can we infer different TCP flavours within transit networks?

Why might we care about this?

Not all algorithms behave fairly!

Yet users expect fair service...

Probably! In rates...

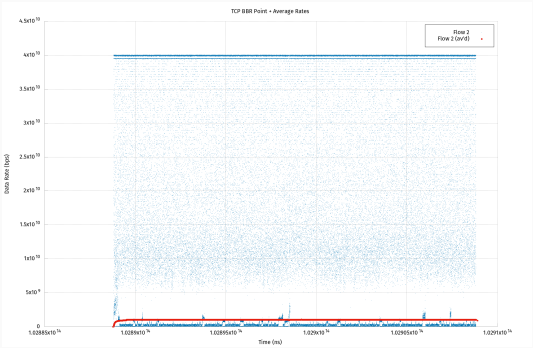


Figure 4: *BBR Rates*

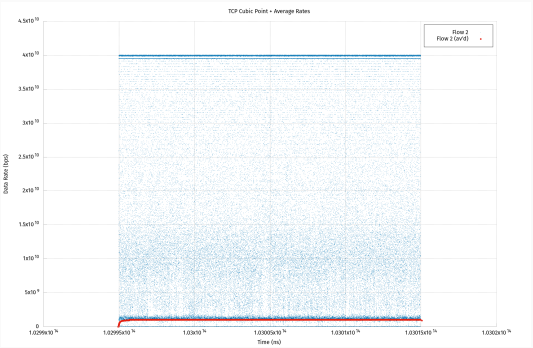


Figure 5: *Cubic Rates*

Probably! In arrival times...

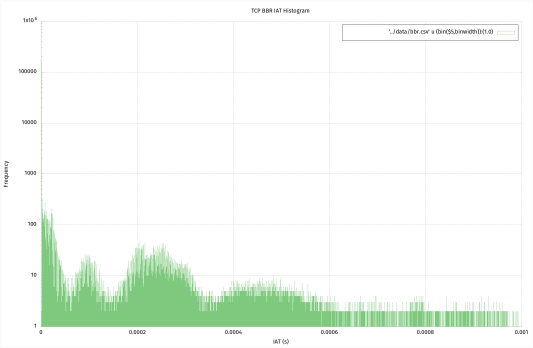


Figure 6: *BBR IATs*

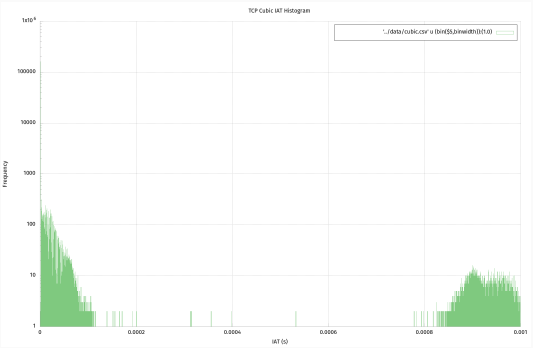


Figure 7: *Cubic IATs*

What about link-limited traffic? (Rates)

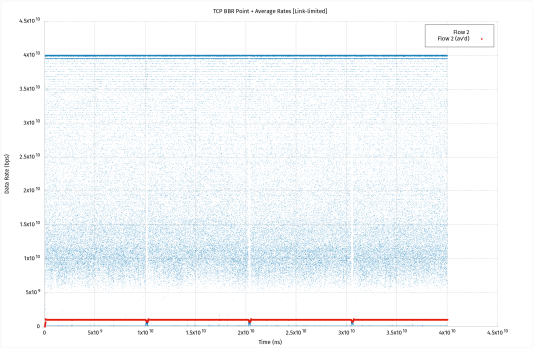


Figure 8: *TC'd BBR Rates*

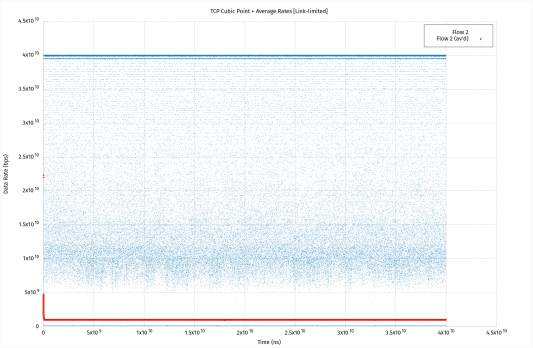


Figure 9: *TC'd Cubic Rates*

The differences are subtler...

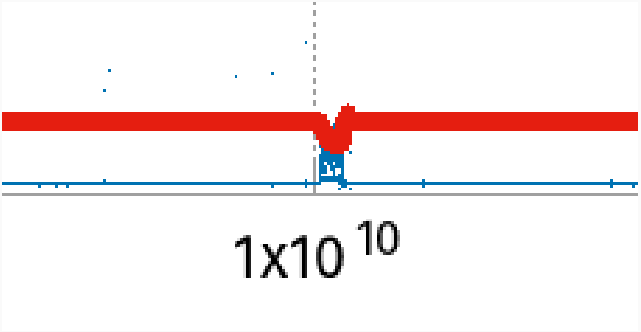


Figure 10: *TC'd Cubic Rates (Zoom)*

What about link-limited traffic? (IATs)

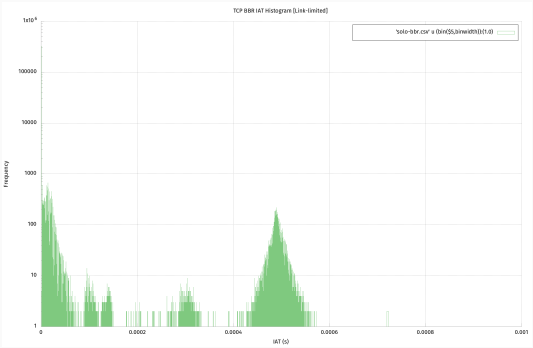


Figure 11: *TC'd BBR IATs*

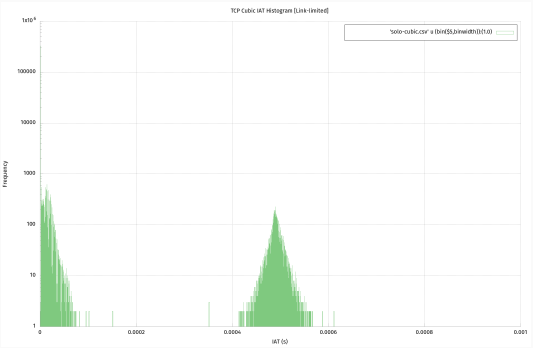


Figure 12: *TC'd Cubic IATs*

Future Work

- Configure *which* analyses are computed.
- IPv6 support.
- Correlating results from SmartNICs[3].
- Microburst detection[4].
- Estimate ‘whole’ RTT from CWnd estimation?
- Programmatically determine TCP flavour—LSTMs, classify by clusters of *dt*s?

[3] Kannan, Joshi and Chan, ‘Precise Time-synchronization in the Data-Plane using Programmable Switching ASICs’.

[4] Chen *et al.*, ‘Catching the Microburst Culprits with Snappy’.

The HighTouch Collector allows high-throughput flow analysis at the edge of ESnet6.

The pipelined design is a core aspect of making this possible.

Importantly, we've seen some interesting flow data, and the insights we can gain from it.

Questions?

k.simpson.1@research.gla.ac.uk

FelixMcFelix https://mcfelix.me

References i

- Chen, Xiaoqi, Shir Landau Feibish, Yaron Koral, Jennifer Rexford and Ori Rottenstreich. 'Catching the Microburst Culprits with Snappy'. In: *Proceedings of the Afternoon Workshop on Self-Driving Networks, SelfDN@SIGCOMM 2018, Budapest, Hungary, August 24, 2018*. ACM, 2018, pp. 22–28. DOI: 10.1145/3229584.3229586. URL: https://doi.org/10.1145/3229584.3229586.
- Ghasemi, Mojgan, Theophilus Benson and Jennifer Rexford. 'Dapper: Data Plane Performance Diagnosis of TCP'. In: *Proceedings of the Symposium on SDN Research, SOSR 2017, Santa Clara, CA, USA, April 3-4, 2017*. ACM, 2017, pp. 61–74. ISBN: 978-1-4503-4947-5. DOI: 10.1145/3050220.3050228. URL: https://doi.org/10.1145/3050220.3050228.

References ii

- Kannan, Pravein Govindan, Raj Joshi and Mun Choon Chan. ‘Precise Time-synchronization in the Data-Plane using Programmable Switching ASICs’. In: *Proceedings of the 2019 ACM Symposium on SDN Research, SOSR 2019, San Jose, CA, USA, April 3-4, 2019*. ACM, 2019, pp. 8–20. ISBN: 978-1-4503-6710-3. DOI: 10.1145/3314148.3314353. URL: https://doi.org/10.1145/3314148.3314353.
- Karn, Phil and Craig Partridge. ‘Improving round-trip time estimates in reliable transport protocols’. In: *Computer Communication Review* 17.5 (1987), pp. 2–7. DOI: 10.1145/55483.55484. URL: https://doi.org/10.1145/55483.55484.